#398

Coherent and Continuous Inference

by

David A. Lane and William D. Sudderth*

University of Minnesota

*Research supported by National Science Foundation Grant MCS8100789.

Abbreviated Title: Coherent and Continuous Inference.

Abstract

If the sampling model is a continuous function of the parameter and if either the parameter space or the observation space is compact, then a coherent inference which is a continuous function of the observation must be the posterior of a proper, countably additive prior.

AMS 1980 subject classifications. 62A15, 60A05

Key words and phrases. Coherence, continuous inference, finitely additive distributions, proper priors.

1. Introduction

An experiment is to be performed, whose possible outcomes are the elements of a set X. The parameter space Θ indexes a family of probability distributions on X; each element θ in Θ corresponds to a possible state of nature, and the distribution p_θ indexed by θ describes the stochastic structure of the experiment, if θ obtains. An inference q is a function, which assigns to every x in X a probability distribution q_x on Θ. An inference might correspond to a system of confidence intervals, a posterior distribution, or a fiducial distribution.

An inference q can be given an operational interpretation, as follows: q_x may be viewed as a conditional odds function, which the inferrer uses to post odds on subsets of Θ after seeing x. Following Freedman and Purves (1969) and Cornfield (1969), the inferrer or his inference q is called <u>coherent</u> if it is impossible for a gambler to devise a system based upon q, which consists of placing a finite number of bets on subsets of Θ after x is observed and which attains an expected payoff greater than some positive constant for every possible state of nature θ. (See Heath and Sudderth (1978) for a precise definition of coherence.)

Heath and Sudderth (1978) showed that coherent inferences correspond to posterior distributions of proper, finitely additive priors on Θ. Now there are technical difficulties involved in employing finitely additive distributions which are not countably additive. For example, such priors may fail to yield posteriors; and even if posteriors exist for a finitely additive prior, there may be no available algorithm, corresponding to Bayes' Theorem, for computing them. The setting in which Heath and Sudderth worked was very general: Θ and X are arbitrary sets, and no compatibility or continuity conditions are imposed upon the p_θ's or the q_x's. Most common inferential problems have much more structure than this. Thus, it is natural to ask: in the kind of problems usually encountered in statistical practice, is every coherent inference available as the posterior calculated from a proper, countably additive prior? The answer to this question is evidently no: for example, if $\Theta = X = R^1$, p_θ is $N(\theta,1)$, and q_x is $N(x,1)$, then the inference q is coherent by [Heath and Sudderth (1978), Example 4.1], but it is not derivable from any proper countably additive prior on Θ.

The main purpose of this paper is to describe a fairly general inferential setting in which all coherent inferences can be obtained as posteriors from proper, countably additive priors. First, suppose both Θ and X are separable metric spaces. Second, suppose each p_θ and each q_x are countably additive. Third, suppose the p_θ's and the q_x's are both weakly continuous. These three conditions are not enough, as the example of the preceding paragraph shows; note, however, that virtually all the classical problems of statistical inference satisfy them. Finally suppose Θ or X is compact: corollary 3.1 asserts that this, together with the three previous conditions, assures that all coherent inferences are posteriors from proper, countably additive priors. Many standard inferential settings satisfy the four conditions. For example, suppose Θ is the closed unit interval, and the experiment consists of one of the usual ways of repeating bernoulli trials (n repetitions, perhaps, or sampling until the m^{th} success): see examples 3.1 and 3.2 for a discussion. Or again, suppose the experiment consists of observing a collection of uniformly truncated survival times, with any standard life-testing parametric model: the uniform truncation gives a compact X.

The next section contains some preliminary material and introduces a notion of consistency of a sampling model with an inference. It turns out that consistency is equivalent to coherence (Proposition 2.1) and is also equivalent to an absence of "strong inconsistency" in the sense of Stone (1976) (Proposition 2.3). The study of consistency may be of independent interest giving, as it does, conditions for the existence of a joint distribution with given conditional distributions.

The major mathematical tool used throughout is the separating hyperplane theorem which was first used in a similar context by Freedman and Purves (1969) and subsequently by Buehler (1965), Heath and Sudderth (1972, 1978), and Quiring (1972).

2. Coherence and consistency.

For any set S, $P(S)$ denotes the collection of finitely additive probability measures defined on all subsets of S. If φ is a bounded, real-valued function defined on S and $\gamma \in P(S)$, then the γ-integral of φ will be written $\gamma(\varphi)$, $\int \varphi d\gamma$, or $\int \varphi(s)\gamma(ds)$.

Let Θ and X be nonempty sets to be regarded as the sets of possible states of nature and possible observations. A sampling model p is a mapping which assigns to each $\theta \in \Theta$ an element p_θ of $P(X)$, and an inference q assigns to each $x \in X$ an element q_x of $P(\Theta)$. Thus p is a conditional distribution on X given Θ and q is a conditional distribution on Θ given X.

Let $r \in P(\Theta \times X)$ and define the marginals π and m of r by

$$\pi(A) = r(A \times X), \quad A \subset \Theta,$$

$$m(B) = r(\Theta \times B), \quad B \subset X.$$

Let $B(\Theta)$ and $B(X)$ be σ-fields of subsets of Θ and X, respectively, and let $B = B(\Theta) \times B(X)$ be the product σ-field.

Roughly speaking, p and q are called consistent if they are the conditional distributions corresponding to some joint distribution r. Here is the precise definition.

Definition. p and q are consistent if there exists $r \in P(\Theta \times X)$ such that, for every bounded, B - measurable function $\varphi : \Theta \times X \to R$,

$$r(\varphi) = \int p_\theta(\varphi_\theta)\pi(d\theta) \tag{2.1}$$

$$= \int q_x(\varphi^x)m(dx).$$

Here, $\varphi_\theta(x) = \varphi(\theta,x)$ and $\varphi^x(\theta) = \varphi(\theta,x)$ for all $(\theta,x) \in \Theta \times X$.
If no such r exists, p and q are inconsistent.

Note: The primary reason for introducing σ-fields is that, very often, p and q are naturally given as countably additive probabilities on some particular σ-field (as when p_θ is $N(\theta,1)$ on the Borels of R^1): it is then reasonable to ask for consistency relative only to the naturally defined parts of these distributions, rather than to the more or less arbitrary extensions to all subsets.

Notice that, if r,p,q satisfy (2.1), then the marginals π and m satisfy

$$\pi(\Psi) = \int q_x(\Psi)\, m(dx) \tag{2.2}$$

$$m(\Psi') = \int p_\theta(\Psi')\, \pi(d\theta)$$

when $\Psi : \Theta \to R$ is bounded, $B(\Theta)$ - measurable and $\Psi' : X \to R$ is bounded, $B(X)$ - measurable.

The proposition below states that coherence as defined in Heath and Sudderth (1978) and consistency are essentially the same concepts.

Proposition 2.1. q is coherent for a given p if and only if p and q are consistent.

Proof: If q is coherent for p, then, by [Heath and Sudderth (1978)], Corollary 1], there is a $\pi \in P(\Theta)$ such that q is the posterior corresponding to π. Take r to be the measure defined by the first equality of (2.1).

Conversely, if p and q are consistent and r is as in the definition of consistency, then q is clearly the posterior of π, the marginal of r on Θ. □

The next proposition gives several conditions equivalent to consistency which will be used in the sequel.

Proposition 2.2. The conditions below are equivalent. (In (b), (c), and (d), $\varphi : \Theta \times X \to R$ is assumed to be bounded and B - measurable.)

(a) p and q are consistent.

(b) For all φ, $\inf_{\theta} \{p_\theta(\varphi_\theta) - \int q_x(\varphi^x) p_\theta(dx)\} \le 0$.

(c) For all φ, $\inf_{x} \{q_x(\varphi^x) - \int p_\theta(\varphi_\theta) q_x(dx)\} \le 0$.

(d) There exist $\pi \in P(\Theta)$, $m \in P(X)$ such that, for all φ ,

$$\int p_\theta(\varphi_\theta)\pi(d\theta) = \int q_x(\varphi^x) m(dx) .$$

Proof: It will be shown that $a \Rightarrow b \Rightarrow d \Rightarrow a$. There is no difficulty in replacing condition b by the symmetric condition c .

$a \Rightarrow b$. Let r be as in the definition of consistency and let π, m be its marginals. Set

$$\Psi(\theta) = p_\theta(\varphi_\theta) - \int q_x(\varphi^x)\, p_\theta(dx) . \tag{2.3}$$

By (2.2) and (2.1),

$$\begin{aligned} \pi(\Psi) &= \int p_\theta(\varphi_\theta)\, \pi(d\theta) - \int q_x(\varphi^x)\, m(dx) \\ &= r(\varphi) - r(\varphi) \\ &= 0 . \end{aligned}$$

Thus, $\inf \Psi$ cannot be positive.

$b \Rightarrow d$. Let C be the collection of all functions Ψ as in (2.3). Because C is convex, it follows from a standard separation theorem [Dunford and Schwartz (1958), p. 417] or more directly from Lemma 1 of Heath and Sudderth (1978), that there is a π in $P(\Theta)$ such that $\pi(\Psi) \ge 0$ for all $\Psi \in C$. But $\Psi \in C \Rightarrow -\Psi \in C$. So $\pi(\Psi) = 0$ for $\Psi \in C$: that is, for all bounded B-measurable θ,

$$\int p_\theta(\varphi_\theta)\pi(d\theta) = \iint q_x(\varphi^x) p_\theta(dx)\pi(d\theta). \tag{*}$$

Now define m by the second equality of (2.2). So (*) is (d).

$d \Rightarrow a$. Define r by the first equality of (2.1) for every bounded $\varphi : \Theta \times X \to R$. By (d), the X-marginal of r is m. That the second equality of (2.1) holds for all φ also follows from (d). $\square$

Following Stone (1976), define p and q to be <u>strongly inconsistent</u> if there is a bounded, B - measurable function $\varphi : \Theta \times X \to R$ such that

$$\inf_{\theta} p_\theta(\varphi_\theta) > \sup_x q_x(\varphi^x) . \tag{2.4}$$

This condition states that the conditional expectations of φ given θ are uniformly larger than those given x.

<u>Proposition</u> 2.3. p and q are strongly inconsistent if and only if p and q are inconsistent.

<u>Proof</u>: Suppose p and q are not strongly inconsistent and let C be the collection of all functions Ψ on $\Theta \times X$ defined by

$$\Psi(\theta,x) = p_\theta(\varphi_\theta) - q_x(\varphi^x)$$

for some bounded, B - measurable $\varphi : \Theta \times X \to R$. Since (2.4) does not hold,

$$\inf_{\theta,x} \Psi(\theta,x) = \inf_{\theta} p_\theta(\varphi_\theta) - \sup_x q_x(\varphi^x) \leq 0 ,$$

for every $\Psi \in C$. By Lemma 1 of Heath and Sudderth (1978), there is a measure $\gamma \in P(\Theta \times X)$ such that $\gamma(\Psi) \geq 0$ for all $\Psi \in C$. But $\Psi \in C \Rightarrow -\Psi \in C$ and so $\gamma(\Psi) = 0$ for all $\Psi \in C$. Thus, for every Ψ,

$$\gamma(\Psi) = \int p_\theta(\varphi_\theta)\pi(d\theta) - \int q_x(\varphi^x)\, m(dx) = 0 ,$$

where π and m are the marginals of γ on Θ and X , respectively. Hence, the second equality of (2.1) holds for all bounded, measurable φ , and r can be defined by the first equality of (2.1) for all bounded φ . This proves that p and q are consistent.

For the converse, notice that (2.4) implies that the second equality of (2.1) fails for every π and m . □

Several interesting examples of strong inconsistency and, therefore, of inconsistency and incoherence are in Stone (1976).

3. Continuous inferences.

For the rest of the paper, Θ and X are assumed to be separable metric spaces with $B(\Theta)$ and $B(X)$ their σ-fields of Borel subsets. Hence, the product σ-field $B = B(\Theta) \times B(X)$ is the σ-field of Borel subsets of $\Theta \times X$ [Parthasarathy (1967), Theorem 1.10]. Let $M(\Theta)$ and $M(X)$ denote the sets of countably additive probability measures on $B(\Theta)$ and $B(X)$, respectively. A measure φ in $P(\Theta)$ (respectively $P(X)$) whose restriction to $B(\Theta)$ ($B(X)$) is countably additive, will, for simplicity, be identified with that restriction.

It will also be assumed from now on that p and q are continuous mappings from Θ to $M(X)$ and X to $M(\Theta)$, respectively, when $M(X)$ and $M(\Theta)$ are given the usual weak topology (as defined in section II. 6 of Parthasarathy (1967)). This assumption of continuity seems quite mild since, to our knowledge, all of the classical likelihoods and inferences satisfy it. Notice that p and q are, in particular, regular conditional distributions, which is equivalent to saying they are Borel measurable mappings.

Two technical lemmas are needed. In the first lemma, the metric d on $\Theta \times X$ is taken to satisfy

$$(3.1) \qquad d\Big((\theta,x)\ ,\ (\theta',x)\Big) = \rho(\theta,\theta')$$

where ρ is the metric on Θ. For example, one could take

$$d\Big((\theta,x)\ ,\ (\theta',x')\Big) = \rho(\theta,\theta') + \tau\,(x,x')$$

where τ is the metric on X.

<u>Lemma</u> 3.1. Let $\varphi : \Theta \times X \to R$ be bounded and uniformly continuous. Then the functions $\theta \to p_\theta\,(\varphi_\theta)$ and $x \to q_x(\varphi^x)$ are continuous too.

<u>Proof</u>: Let $\theta_n \to \theta$. It suffices to show $p_{\theta_n}(\varphi_{\theta_n}) \to p_\theta(\varphi_\theta)$.

By (3.1), $(\theta_n,x) \to (\theta,x)$ uniformly in x and, hence, $\varphi(\theta_n,x) \to \varphi(\theta,x)$ uniformly in x.

Now write

$$p_{\theta_n}(\varphi_{\theta_n}) - p_\theta(\varphi_\theta) = \left[\int\varphi(\theta_n,x)\ p_{\theta_n}(dx) - \int\varphi(\theta,x)\ p_{\theta_n}(dx)\right]$$
$$+\left[\int\varphi(\theta,x)p_{\theta_n}(dx) - \int\varphi(\theta,x)\ p_\theta\ (dx)\right].$$

The first bracketed expression on the right converges to zero because $\varphi(\theta_n,x) \to \varphi(\theta,x)$ uniformly in x; the second expression converges to zero because $p_{\theta_n} \to p_\theta$ weakly. □

Let $\mathcal{C}(\Theta)$ be the collection of bounded, real-valued continuous functions defined on the metric space Θ, and give $\mathcal{C}(\Theta)$ its sup norm topology.

<u>Lemma</u> 3.2 Assume Θ is compact and let C be a convex subset of $\mathcal{C}(\Theta)$. Then the following are equivalent:

(i) $\inf \Psi \leq 0$ for all $\Psi \in C$.

(ii) There exists a π in $M(\Theta)$ such that $\pi(\Psi) \leq 0$ for all $\Psi \in C$.

<u>Proof</u>: The proof is similar to that of Lemma 1 in Heath and Sudderth (1978) and uses a separating hyperplane theorem [Dunford and Schwartz (1958), p. 417] together with Riesz's Theorem [Parthasarathy (1967), Theorem II 5.8] which characterizes the nonnegative, normed linear functionals on (Θ) as the elements of $M(\Theta)$. □

<u>Proposition</u> 3.1. If p and q are continuous and Θ is compact, then the following are equivalent.

(a) p and q are consistent.

(b) For every bounded, uniformly continuous function

$\varphi : \Theta \times X \to R$,

$$\inf_{\theta} \{p_\theta(\varphi_\theta) - \int q_x(\varphi^x) p_\theta(dx)\} \leq 0 .$$

(c) There exist $\pi \in M(\Theta)$, $m \in M(X)$ such that, for every bounded, Borel measurable function

$\varphi : \Theta \times X \to R$,

$$\int p_\theta(\varphi_\theta)\, \pi(d\theta) = \int q_x(\varphi^x)\, m(dx) . \tag{3.2}$$

Furthermore, (a) and (c) remain equivalent if the hypothesis that Θ is compact is replaced by one that X is compact.

Proof: That $a \Rightarrow b$ and $c \Rightarrow a$ follows from Proposition 2.2.

$b \Rightarrow c$. The proof is similar to the proof that $b \Rightarrow d$ in Proposition 2.2. Use Lemma 3.1 to show that, for bounded uniformly continuous φ, the Ψ defined by (2.3) is in $C(\Theta)$ and then use Lemma 3.2 to find $\pi \in M(\Theta)$ so that $\pi(\Psi) \leq 0$ and, hence, $\pi(\Psi) = 0$ for all Ψ corresponding to a bounded, uniformly continuous φ. If m is defined by (2.2), then (3.2) holds for all such φ. But two countably additive probability measures which agree on the bounded, uniformly continuous functions must agree on all bounded Borel functions [Theorem II.5.9 of Parthasarathy (1967)].

The final assertion of the proposition is obvious because of the symmetry of conditions (a) and (c) in θ and x. □

Corollary 3.1. Suppose p and q are continuous and either Θ or X is compact. Then q is coherent for p if and only if q is the posterior of a prior $\pi \in M(\Theta)$ (i.e. a proper, countably additive prior).

Proof: Suppose q is coherent for p. By Proposition 2.1, p and q are consistent. By Proposition 3.1 (part(c)), q is the posterior for a $\pi \in M(\Theta)$.

The opposite implication is immediate from Proposition 2.1. □

Here are two examples, both concerning inference about the probability of success in Bernoulli trials, which illustrate the application of Corollary 3.1.

Example 3.1: Suppose x represents the number of successes in n Bernoulli trials, and θ is the probability of success in a single trial. Thus

$$X = \{0,1,\ldots,n\} ,$$

and

$$p_\theta\{x\} = \ell_x(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} .$$

For each x, ℓ_x is a polynomial of degree n, and so p is certainly weakly continuous. Since X has the discrete topology, any inference q is weakly continuous. And since X is finite, it is compact, so Corollary 3.1 applies.

Suppose Θ is the open interval $(0,1)$: this choice of parameter space reflects the judgement that the experiment is not purely deterministic, and both success and failure may be observed. With this choice of parameter space, each x in X receives positive probability under each θ in Θ. By Corollary 3.1 and Bayes' Theorem, then, every coherent inference for this problem can be expressed as

$$(3.2) \qquad q_x(d\theta) = \frac{\ell_x(\theta)\mu(d\theta)}{\int \ell_x d\mu}$$

for some countably additive probability measure μ on Θ.

A consequence of this result is that improper Bayesian inferences are incoherent for this problem. Here, improper Bayesian inference refers to the following procedure: select a countably additive measure μ on the Borel subsets of Θ with infinite total mass; define q_x by Bayes' formula (3.2) when the denominator is finite and define q_x to be an arbitrary element of $M(\Theta)$ otherwise. Such a q cannot be the inference corresponding to any proper, countably additive ν unless $\int \ell_x d\mu$ is infinite for all x. For suppose that integral is finite. Then q_x is given by (3.2) and cannot be the measure

$$\frac{\ell_x(\theta)\nu(d\theta)}{\int \ell_x d\nu}$$

because, for example, ℓ_x^{-1} has finite integral for this measure but not for q_x. By Corollary 3.1 then, this improper inference must be incoherent. A common choice of μ is

$$(3.3) \qquad \frac{1}{\theta(1-\theta)}\, d\theta$$

see, for example Jeffreys (1961) or Jaynes (1968). Now an <u>actual</u> improper Bayesian, like Jeffreys, would not, if he chose the prior in (3.3), select any measure in $M(\Theta)$ as his inference if x were 0 or n. Rather, he would declare an <u>improper</u> posterior. The point we are making here is that he <u>could not</u> modify the formal posterior computed from such improper priors as those in (3.3) by selecting q_0 and q_n from $M(\Theta)$, so as to produce a coherent inference.

<u>Example 3.2</u>: x and p_θ the same as in example 3.1; but Θ is the closed interval $[0,1]$. Suppose the improper Bayesian selects (3.3) as his "prior",

and uses its formal posterior, modified as follows:

$$q_x(\theta) = \begin{cases} c(x)\theta^{x-1}(1-\theta)^{n-x-1} & \text{for} \quad 1 \leq x \leq n-1 \\ \delta_0 & \text{for} \quad x = 0 \\ \delta_1 & \text{for} \quad x = n \, . \end{cases}$$

This inference is now coherent, as it is a posterior for the proper, countably additive prior

$$\mu(d\theta) = \tfrac{1}{2}(\delta_0 + \delta_1),$$

or _any_ prior concentrated on $\{0,1\}$.

This example illustrates the fact that coherence may provide a necessary condition for reasonable inference, but is not sufficient. The inferrer with prior concentrated on $\{0,1\}$ is free to be stupid should at least one success and one failure occur.

If neither θ nor X is compact, but the other hypotheses of this section remain in force, then a continuous q is coherent for a continuous p if and only if q is the posterior of a finitely additive π which is regular on the field Σ generated by the closed subsets of Θ. (The proof is similar to that given above. However, the crucial separation argument of Lemma 3.2 uses the fact that the nonnegative, normed linear functionals on $C(\Theta)$ are the regular π's [Dunford and Schwartz (1958), p. 262]). Such a π, i.e. one regular on the field Σ, is countably additive when restricted to subsets of a compact set. Thus any failure of π to be countably additive must occur on unbounded sets and is linked to the tail behavior of π.

One of the hypotheses of this section, which has not been emphasized, is that all of the measures q_x are countably additive. If, instead, they are allowed to be finitely additive and if the space $P(\Theta)$ is given an appropriate topology, a purely finitely additive prior π can lead to a continuous q. Here is a trivial example: let $\Theta = [0,1]$, and suppose Υ_0 is a diffuse point mass at 0, and Υ_1 is a diffuse point mass at 1. (That is, $\Upsilon_0(A) = 1$ for A any interval containing 0, but $\Upsilon_0\{0\} = 0$; and similarly for Υ_1.) Let $\pi = \frac{1}{2}\Upsilon_0 + \frac{1}{2}\Upsilon_1$. Now, suppose $X = \{0,1\}$, and $p_\theta(\{0\}) = 1 - p_\theta(\{1\}) = \theta$. Finally, suppose $q_0 = \Upsilon_0$ and $q_1 = \Upsilon_1$. Trivially, q is a continuous inference. Also, q is a posterior for the purely finitely additive prior π, but q is not the posterior for any countably additive prior.

References

Buehler, R.J. (1976). Coherent preferences. Ann. Statist. 4 1051 - 1064.

Cornfield, J. (1969). The Bayesian outlook and its applications. Biometrics 25 617-658.

Dawid, A.P. and Stone, M. (1973). Expectation consistency and generalized Bayes inference. Ann. Statist. 1 478-485.

Dunford, Nelson and Schwartz, Jacob (1958). Linear Operators, Part I: General Theory. Interscience, New York.

Freedman, David and Purves, Roger (1969). Bayes' method for bookies. Ann. Math. Statist. 40 1177-1186.

Heath, David and Sudderth, William (1972). On a theorem of de Finetti, oddsmaking and game theory. Ann. Math. Statist. 43 2072-2077.

Heath, David and Sudderth, William (1978). On finitely additive priors, coherence, and extended admissibility. Ann. Statist. 6 333-345.

Jeffreys, H. (1961). Theory of Probability (Third Edition), Oxford University Press, London.

Jaynes, E. (1968). "Prior Probabilities," IEEE Transactions on Systems Science and Cybernetics SSC-4, 227-241.

Parthasarathy, K.R. (1967). Probability Measures on Metric Spaces. Academic Press, New York.

Quiring, John N. (1972). Coherent odds and Bayesian odds. Ph D thesis and Technical Report No. 179, University of Minnesota School of Statistics, Minneapolis.

Stone, M. (1976). Strong inconsistency from uniform priors. J. Amer. Statist. Assoc. 71 114-116.